Progress Report: ONR Award 431-0857A
"Restricted Cycle Problems with Applications"

Principal Investigator: Steven S. Skiena
Department of Computer Science
State University of New York
at Stony Brook

October 14, 1997

1 Results

This progress report covers the first six months of funding on this grant. This grant covers two distinct lines of work: (1) research on the algorithmic complexity of path and cycle problems in graphs, and (2) implementations of graph algorithms as part of large-scale environments for combinatorial computing. As detailed below, substantial progress has been made in both areas since the commencement of funding.

2 Restricted Path and Cycle Problems in Graphs

- Based on our theory of triangular strip construction for triangulated models commonly used in computer graphics, and implemented our algorithms in a program *Stripe*, our code for triangular strip construction for triangulated models, continues to grow in popularity in the computer graphics community. Over 400 industrial and educational sites have downloaded *Stripe* from our WWW site, *http://www.cs.sunysb.edu/~skiena*, and been rewarded by speeding their typical rendering times by 10-20%.

  Military users to date include: neiderer@arl.mil, garrett@alhra.af.mil, and reif@rdvax.ntsc.navy.m In particular, Jacki L. Garrett reports "I'm hoping that Stripe will help me turn inefficient models (from a variety of sources) into relatively efficient real-time models for use in flight simulators. Thanks for making Stripe accessible!"

  Stripe is based on our theory [9] of finding a special type of Hamiltonian cycle in the dual graph of the triangulation.

- Using information-theoretic constraints, we [6] have developed a system for removing the ambiguity in English text typed on heavily overloaded keyboards, achieving a reconstruction accuracy of over 99%. This can be applied to variety of user interface applications, including data input for virtual reality (VR) systems, eyetracking systems

for the handicapped, and e-mail without a terminal. The key algorithmic problem is finding the shortest path in a graph.

- In [3], we developed a new algorithm for efficient pattern matching in run-length encoded strings, which are commonly used in image processing applications such as graphics morphing and OCR [7]. In particular, we find the longest common subsequence in two strings in time $O(st \log(s+t))$, where s and t are the compressed run-lengths of the two strings.

- In collaboration with biologists at Brookhaven National Laboratory, we have developed *STROLL* [5, 4], a state-of-the-art fragment assembler for genome-level "shotgun sequencing". The Brookhaven group used *STROLL* in their successful project to sequence the 1 megabase genome of the bacteria Borrelia Burgdorferi, which causes Lyme disease. The problem of fragment assembly is akin to finding the longest path in a graph defined by the overlaps of strings.

3 Environments for Combinatorial Computing

- We have finished and released the initial version of the *Stony Brook Algorithm Repository*, a collection of implementations of combinatorial algorithms available on the WWW at http://www.cs.sunysb.edu/~algorith. Our site has already been accessed by over 11,000 people in it prerelease version, including almost 150 Navy and military sites.

- I have completed the final page proofs of my book [1], "The Algorithm Design Manual", to be published by Springer-Verlag on October 30, 1997. This book is integrated with the Stony Brook Algorithm Repository, described.

- A paper describing *LINK* [2], our environment for combinatorial computing, was presented at the conference Graph Drawing '97. We have had preliminary discussions with Wolfram Research, Inc. concerning the design of a graph algorithm kernal within *Mathematica*

As to the impact on education human and resources, over the period of ONR funding I have graduated one Ph.D students (Ting Chen), who has just taken a research position at Harvard Medical School. I am currently advising two additional Ph.D students (Barry Cohen and Pavel Sumazin) and co-advising two others (Francine Evans and George Sazaklis).

Among other accomplishments, I have been named to the Program committee for the *Fourteenth ACM Symposium on Computational Geometry* (applied track), Minneapolis MN, June 7-9, 1998.

A list of recent publications by the PI acknowledging ONR grant support follows.

References

[1] S. Skiena, *The Algorithm Design Manual*, Telos/Springer-Verlag, New York, 1997.

[2] Graph Drawing and Manipulation with LINK, (with J. Berry, N. Dean, M. Goldberg, and G. Shannon). *Proc. Graph Drawing '97*, Lecture Notes in Computer Science, Springer-Verlag, September 1997.

[3] Matching for Run-Length Encoded Strings (with A. Apostolico and G. Landau). *Sequences '97*, Positano Italy, June 11-13, 1997. Also, submitted to *Journal of Complexity*, special issue.

[4] Trie-Based Data Structures for Sequence Assembly (with T. Chen). Proc. *Combinatorical Pattern Matching (CPM '97)*.

[5] How Good is Genome-Level Fragment Assembly? (with T. Chen), Submitted to *RECOMB '98*.

[6] Resolving Ambiguity in Overloaded Keyboards (with F. Evans and A. Varshney), submitted to *ACM Transactions on Human-Computer Interfaces*.

[7] Geometric Decision Trees for Optical Character Recognition (with G. Sazaklis, E. Arkin, and J. Mitchell). *13th ACM Symp. Computational Geometry*, short communications, applied track. June 1997.

[8] Efficiently Partitioning Arrays (with S. Khanna and M. Muthukrishnan). Proc. *ICALP '97*.

[9] Optimizing Triangle Strips for Fast Rendering (F. Evans, S. Skiena, and A. Varshney). Submitted to *ACM Transactions on Graphics*

[10] Enabling Virtual Reality for Large-Scale Mechanical CAD Datasets (with A. Varshney, J. El-Sana, F. Evans, L. Darsa, and B. Costa). *Proc. 1997 ASME Design Engineering Technical Conferences* (DETC'97), Sacremento CA, September 14-17, 1997.